DOCUMENT RESUME

ED 081 783 TM 003 143

TITLE	Multiple Choice Questions: A Close Look.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	73
NOTE	17p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Cognitive Tests; *Multiple Choice Tests; *Objective Tests; *Statistical Analysis; *Test Construction; *Test Results

ABSTRACT

A sample of 22 multiple-choice questions is provided to point up the potentialities of this type question, as well as to dispel the myth that "objective" tests require no thought, insight, or understanding. The questions have been taken from a variety of tests constructed by Educational Testing Service, and are presented as they appeared in the original tests, with the directions, when given, as they were given to the students. Each question is followed by a statistical analysis of its performance and by a brief discussion of the thought processes involved in formulating the questions and arriving at the correct answer. (DB)

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MULTIPLE CHOICE QUESTIONS:

A Close Look

EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY

ED 081783

Copyright © 1963, 1973 by Educational Testing Service. All rights reserved.

TM 003 143

The purpose of this close look at a group of multiple-choice questions is to dispel a myth: the myth of the multiple-choice question as a superficial exercise—one that requires little thought, less insight, and no understanding. Like other myths, this one may be based on a shadowy memory of the past, but it bears little relation to present reality. "What is often overlooked . . . ," writes Jerome Bruner in *The Process of Education*, "is that examinations can also be allies in the battle to improve curricula and teaching. Whether an examination is of the 'objective' type involving multiple choices or of the essay type, it can be devised so as to emphasize an understanding of the broad principles of a subject. Indeed, even when one examines detailed knowledge, it can be done in such a way as to require understanding by the student of the connectedness between specific facts."

Educational testing in its present form and prevalence is a response to the needs of contemporary American education. Tests of scholastic aptitude and achievement seek to measure certain abilities that educators want evaluated. At Educational Testing Service, this measurement is accomplished by leaving to the educators the responsibility for determining the objectives of each test and by intimately involving them in the actual construction of test questions. The great majority of tests published by ETS are prepared by committees of school and college teachers. The committees are selected to represent different institutions, regions, and points of view, as well as the breadth and depth of scholarship needed to prepare a test in a particular subject. In a typical year, over 500 individuals serve on such committees.

If a test is to have relevance to education, a test committee, before writing a single question, must consider the goals of instruction that are most important to measure. Most often this step results in the preparation of a written statement or outline of what is to be tested. Without such forethought a test may measure very precisely some very unimportant things. If tests are to be educationally useful, the test questions must demand more from the student than rote recall. The preparation of written test specifications gives each committee the opportunity to identify the knowledge and the skills that ought to be measured by the test questions, and to lay specific plans for the preparation of such questions.

The actual creation of the test questions is itself a crucial, arduous process. It is crucial because the loftiest of objectives are meaningless if the questions in a test do not require the students taking the test to demonstrate attainment of those objectives. It is arduous because of the several important steps that must be successfully completed in the writing of the questions.

First, it is essential that the author of a test question identify and define the skill or knowledge he wishes to measure. He must satisfy himself that, among the many questions that may be asked, a particular question is worth asking. Next, he must devise a way of asking the question that will insure that the correct answer is not identified in some superficial manner. As may be apparent in the examples which follow, this consideration requires precision and artistry in the writing of test questions. A well conceived and constructed multiple-choice question should require a student to select, weigh, and apply what he knows in order to answer the question correctly.

Also, the author of the question must word the question and the answer choices so that those who indeed know what is being asked can answer correctly and those who are less knowing will, in effect, be defeated by their own ignorance. The art of writing a multiple-choice question requires, of course, a knowledge of the subject and a knowledge of how the subject is taught at the educational level for which the question is intended. Moreover, the writer of the question should be familiar not only with what students ought to have learned, but also with what they ought not to have learned. That is, he should be able to construct each question in such a way that students who have misinformation or misconceptions will be differentiated from those who are knowledgeable. Then each answer choice in a multiple-choice question will contribute to differentiating between competent and less-than-competent students. Statistical analyses of the sample questions that follow will show how this differentiating power of a question depends on the extent to which the wrong answers attract the less competent students. Questions can discriminate at different levels of ability. In effect, a question's difficulty and discriminating power depend on the sophistication required to distinguish among the answer choices.

To create questions that are important, appropriate, accurate, and discriminating requires the imaginative efforts of well-trained, experienced individuals working in concert. Typically, a test committee assumes responsibility not only for planning a test but for writing and reviewing the test questions as well. Usually each committee member writes a certain number of questions and reviews those submitted by his colleagues. An ETS test development specialist assists the committee in planning the questions, offers advice on the construction of the questions, and reviews the questions as a technical editor. Finally, the committee selects from among all the questions written those that it deems of sufficiently high quality to use in its test.

After passing the test of expert judgment, questions must then pass a second test—the test of use. Is the question easy—perhaps too easy—or difficult—maybe too difficult? Does it differentiate between able and less able students? Is it clear and unambiguous? Even the most competent teachers and scholars may occasionally misjudge such matters. If a test is to provide reliable measurement, possible sources of error are best identified and eliminated in advance. Therefore, whenever feasible, questions are tried out experimentally before inclusion in a test. From such "pretesting" considerable information of value is derived. Each question in this booklet is followed by a statistical analysis that summarizes the kind of information obtained

FILMED FROM BEST AVAILABLE COPY

about the performance of those students tested with the question *before* its actual use in a test as well as after its actual use.

The difficulty of each question may be readily determined. Difficulty is represented by the percentage of students who answer a question correctly; the higher the percentage correct, the easier the question. Similarly, the attractiveness of each answer choice may be determined by the number of students who select it.

The power of a question to differentiate among students of different academic abilities is a more complex matter. To determine this, it is necessary to have for each student a measure of the ability that is being tested by the question; this measure serves as a criterion. As circular as it may at first seem, one such readily available criterion is the student's score on the total test of which the question is a part. The validity of such a technique has been confirmed by comparing the results of analysis against an internal criterion with those obtained with a less readily available independent criterion. In essence, although a committee of experts may err in the formulation of a particular question, we may assume that its judgment of what constitutes achievement in an area is generally valid. Hence the relation between students' success on a question and success on the entire test of which the question is a part is an appropriate measure of the power of the question to discriminate between more able and less able students.

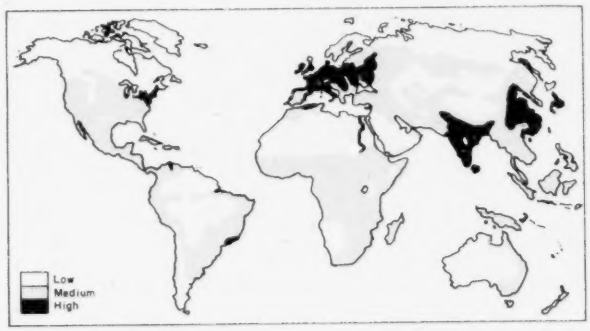
But since much of the power of a multiple-choice question depends on the nature of the answer choices, obtaining measures solely of the question's over-all performance is not sufficient. If the strengths and weaknesses of a question are to be diagnosed, knowledge of the performance of each of its strategic parts must also be obtained. Therefore, for each question an analysis is made of the ability of the group of students selecting each answer choice. It seems reasonable to expect that, in general, those who are more able should select the correct response to a question and that, in general, those who are less able should more frequently find wrong answers attractive. To analyze the performance of each question, students may be divided (as illustrated in the following pages) into equal fifths based on their performance on a relevant criterion—usually total test score. When the answer choices in a question are effective, more students in the highest fifth than in the lowest fifth will select the correct response, and more in the lowest fifth than in the highest fifth will select the wrong answers. Also, the number of students selecting the correct response will generally increase in each higher fifth, while the number choosing the wrong responses will increase in each lower fifth. Information of this kind may be used by a test committee to select for its test only those questions which have proven to discriminate appropriately, or which, given the information, may be revised to make them sufficiently discriminating. The questions appearing in this booklet have all passed through such a screening.

Despite all these precautions, not every question prepared by the methods described turns out to be a good one. Even after the most careful scrutiny of an expert committee as well as a painstaking experimental screening, some do not measure adequately an important instructional goal, and some may contain errors of fact or interpretation. These, however, are the rare exceptions.

The questions which appear on the following pages have been taken from a variety of tests constructed by Educational Testing Service. Many similar questions are now in use. Those presented here are as they appeared in the original tests, with the directions, when given, as they were given to the students. Each question is followed by a statistical analysis of its performance and by a brief discussion of the thought processes involved in formulating the question and in arriving at the correct answer.

Study of this representative sample of multiple-choice questions should lead to a clearer understanding of their potentialities, and should help to dispel the myth that "objective" tests require no thought, insight, or understanding.

Question 1



The shading on the above map is used to indicate

(A) population density

(B) percentage of total labor force in agriculture

(C) per capita income

(D) death rate per thousand of pop ion

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	1				
*A	45	52	53	58	58
B	6	1	2	1	1
C	6	5	1		1
D	2	2	4	1	
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 89%
Correlation between success on this question and total score on test .47

*Correct answer

In many of the multiple-choice questions included in tests in the social sciences, an attempt is made to require the student to make use of his general background of knowledge in the interpretation of materials. Thus, this question does not simply ask: What areas of the world have the highest population densities? Rather, it presents a novel situation in which the student must infer that, of the choices offered, only population density provides a plausible explanation of the shadings on the map.

This question was answered correctly by almost 90 per cent of a group of college seniors. However, although large numbers of lower ability students were successful on it, the question did differentiate between some of the very poorest students and the rest of those tested.

An examination of the map clearly shows that choice (A), population density, is the proper response. The darkest shading, which according to the map's legend indicates the highest degree of whatever the shading represents, covers such high population density areas as the northeastern part of the United States, a large part of Europe, the Nile valley, India, Japan, and Eastern China. If this were not a sufficient clue, the areas with the lightest shading include such underpopulated areas as the Arctic regions, tropical South America, the Sahara and Arabian deserts, and most of Australia.

Choice (B), the percentage of total labor force in agriculture, while possibly attractive if only India and China are examined, is clearly incorrect when applied to the Northeastern United States. This choice was taken by 10 per cent of the least able group, but was not attractive to any of the other groups.

Choice (C), per capita income, attracted moderately those in the lowest two-fifths of the group. Per capita income could be plausible only if the student's analysis of the map took in solely the dark shading in the United States and Western Europe, and even there it is not entirely correct, but the dark shading in China and India could certainly not indicate high per capita income. The reverse observation is true of choice (D), death rate per thousand of population, since the latter might be expected to be high for India and China, but low for the United States and Europe. Only nine students selected this choice.

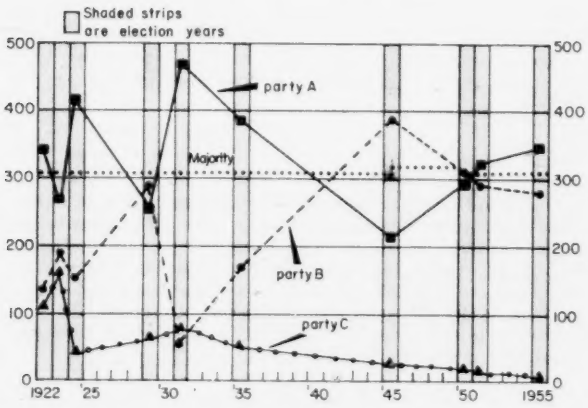
Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	8	1	3	2	
A	5	4	5	4	2
B	25	28	14	11	3
C	6	5	6	5	6
*D	16	22	32	38	49
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 52%
Correlation between success on this question and total score on test ..... .49

*Correct answer

Question 2



The graph above represents the political composition from 1922 to 1955 of which of the following?

(A) German Bundestag

(B) French National Assembly

(C) Italian Chamber of Deputies

(D) British House of Commons

To answer this question correctly, the student must be able to do several things. First, he must be able to read the graph. Then, using the information he can infer from it, he must interpret it in the light of his knowledge of European history and government from 1922 to 1955 and draw a conclusion concerning which legislative body may properly be so depicted. In such a process, it is possible for different students to make use of different information to arrive at the correct answer.

In examining the graph, the student should note that the party system shown is essentially a two-party one, although there is a third party that, for most of the period shown, decreased in representation. He may also note the years in which elections were held, the years that Party A received majorities, the fact that Party B did not receive a majority until 1945, and the fact that in the elections of 1923 and 1929 neither party received a majority.

In considering the first of the four possible answers, the German Bundestag, the student should recognize that during the period through 1932, that of the Weimar Republic, no single party in Germany was able to attain a majority, partly because of the multiplicity of parties. After Hitler came to power in 1933, parties other than the Nazi Party were outlawed. These facts do not fit the graph, as most of the students recognized.

The second possible answer, the French National Assembly, contained far more than three parties both before and after World War II. In spite of this, many students, especially those in the lowest two-fifths, found this choice attractive. Only three of the highest group selected it, however. The Italian Chamber of Deputies should also be rejected as a possible answer because, after Mussolini came to power in 1924, it became less and less important until, in 1938, it was superseded by a Chamber of Fasci and Corporations and political parties were suppressed. This choice attracted only five or six students in each of the fifths.

Over half of the group of college students who attempted this question selected the correct response, the British House of Commons. Party A, on the graph, corresponds to the Conservative Party, Party B to the Labor Party, and Party C to the Liberal Party. The large Conservative majorities of 1924 and 1931 are clearly shown, as is the attainment of a majority by the Laborites in 1945.

Question 3

"In a flash it came upon me that there was the reason for advancing poverty with advancing wealth. With the growth of population, land grows in value, and the men who work it must pay more for the privilege. In allowing one man to own the land on which and from which other men live, we have made them his bondsmen in a degree which increases as material progress goes on. This is the subtle alchemy that in ways they do not realize is extracting from the masses in every civilized country the fruits of their weary toil."

The person most likely to have written these words is

(A) John Jacob Astor

(B) William Jennings Bryan

(C) Thorstein Veblen

(D) Lincoln Steffens

(E) Henry George

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	32	22	15	16	4
A	5	6	2	1	
B	15	13	22	10	9
C	7	10	13	12	16
D	8	8	9	8	4
*E	6	15	13	27	41
Total	73	74	74	74	74

Per cent of total group of 369 students answering correctly 28%
Correlation between success on this question and total score on test . . . 47

(Discrepancies in total numbers in each fifth are caused by drop-out of students not completing the test.)

*Correct answer

This question presents the student with a quotation and asks him to identify the person most likely to have made the statement. Note that the student is *not* asked to, or expected to, recognize the statement from memory. Instead, he is expected to read the excerpt carefully, to evaluate it in terms of his knowledge and understanding of American intellectual history, and then to select from the five names listed the person to whom the statement might most reasonably be attributed.

Some interesting results may be seen by examining the statistics obtained when the question was administered in a college-level American history test to a group of very able high school seniors. Because the question was a difficult one for the group, many students chose not to answer it at all rather than risk being penalized for an incorrect guess. This helped to differentiate between the able students and the less able ones. Whereas 32 of the students in the lowest fifth omitted the question, only four of the students in the top fifth did so.

Choice (A) served to attract only a few of the less able students. None of the top group thought that the statement could have been made by John Jacob Astor. An entrepreneur of the early nineteenth century who built up a vast fur trading empire in the Pacific Northwest, Astor would most likely be associated with the opening up of the Oregon country. In this huge, unsettled region where land was then readily available, he would hardly have been concerned with the particular economic problem discussed in the quotation. Moreover, anyone who recognizes the name has no reason to associate it with reform.

Choice (B) is a more sophisticated wrong answer than is (A). The students who selected William Jennings Bryan as the probable author of the passage were undoubtedly aware of Bryan's reputation as "The Great Commoner"—the self-styled defender of the people against the "dictatorship" of Wall Street, their champion in the campaign for free coinage of silver. Although the last sentence of the quotation might be slightly reminiscent of Bryan's style, the particular ideas are not those exploited by an orator who made free silver a "cause célèbre." This answer, while attractive to some of the students in the lower two-fifths, was most popular with those in the middle fifth. This may be partly explained by its greater sophistication, but the fact that most of the lower group had already chosen to omit the question probably accounts for the failure of that group to choose (B).

Choice (C), Thorstein Veblen, served to separate the better students from the very best. Few of the poorest students chose Veblen, partly because of the reasons mentioned above, and possibly also because they may not have been familiar with Veblen. The brighter students may have been more familiar with his name and ideas or at least with the fact that he was a critic of the American economic system at the turn of the century. They did not know enough, however, to realize that Veblen was not concerned with the monopoly of land as a cause of poverty.

Very few of the top group were attracted to choice (D), Lincoln Steffens. Concerned primarily with the problems of the cities, Steffens is perhaps best known for his exposure of municipal corruption.

Henry George, the correct response, was selected by over half of those in the highest fifth, but only six of the lowest group chose this correct answer. Altogether only 28 per cent of the total group could answer the question correctly, for a thoughtful understanding of the ideas of Henry George was required. The statement did not contain any of the catchwords usually superficially connected with George, such as "single tax" or "unearned increment."

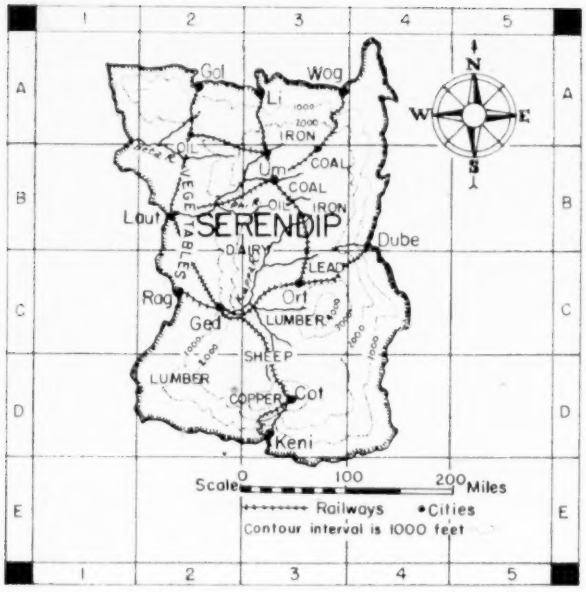
Question 4

In the following questions you are asked to make inferences from the data which are given you on the map of the imaginary country, Serendip. The answers in most instances must be probabilities rather than certainties. The relative size of towns and cities is not shown. To assist you in the location of the places mentioned in the questions, the map is divided into squares lettered vertically from A to E and numbered horizontally from 1 to 5.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Dorothy Urban

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."



Which of the following cities would be the best location for a steel mill?

(A) Li (3A)

(B) Um (3B)

(C) Cot (3D)

(D) Dube (4B)

Statistical Analysis

RESPONSES	Students Classified by Total Test Score	
	LOWEST 27%	HIGHEST 27%
Omit	8	
A	10	2
*B	40	84
C	4	1
D	9	6
Total	71	93

Per cent of total group of 370 students answering correctly 75%
Correlation between success on this question and total score on test . . .43

(A somewhat different form of analysis was used for this test. Discrepancies in total numbers in each group are caused by drop-out of students not completing the test.)

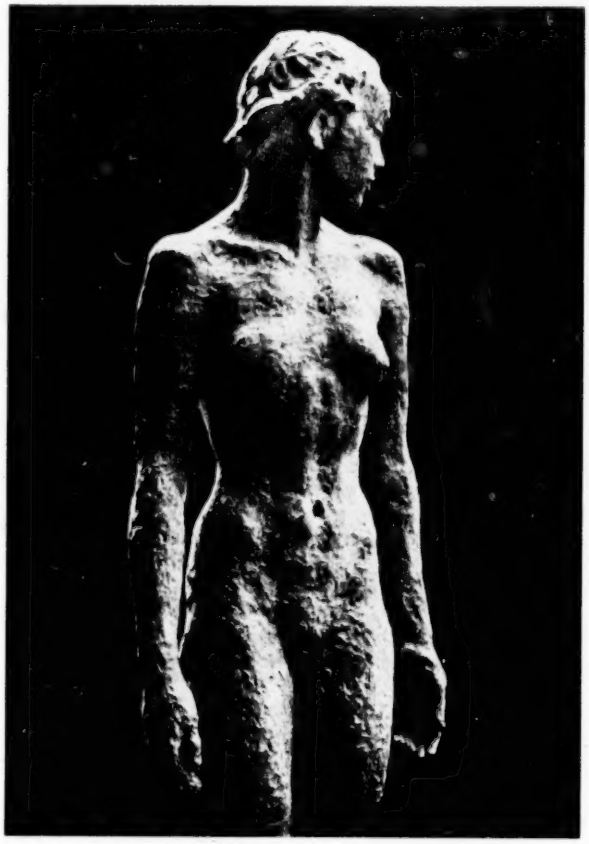
*Correct answer

A map of an imaginary country, such as that shown above, offers numerous possibilities for questions which measure important understandings. One could ask several questions requiring an understanding of the symbols used on the map. To determine student comprehension of the meaning of contour lines, for example, one might ask which railroad has the steepest grades to climb. Similar questions can be developed which require knowledge of the factors influencing population distribution, economic activities and so on.

The question reproduced beneath the map requires knowledge of the natural resources used in producing steel and an awareness of the importance of transportation facilities in bringing these resources together. It was part of a general achievement test given to high school seniors.

The student who knows that iron is the basic raw material of steel and that coal commonly provides the necessary source of heat would proceed to locate deposits of these resources in relation to the cities listed in the question. He would be able to eliminate Cot immediately, since there is no iron or coal in its vicinity, although Cot might be an attractive choice to students who mistakenly think that copper is a basic ingredient of steel. Both Li and Dube are located reasonably near supplies of iron, and therefore might be attractive choices. Um, however, is the more clearly "correct" response, because not only are deposits of iron and coal nearby, but they are more readily transportable by direct railroad routes.

Question 5



In which of the following centuries was the piece of sculpture shown above most probably produced?

(A) The fifth century B.C.

(B) The fourteenth century A.D.

(C) The sixteenth century A.D.

(D) The eighteenth century A.D.

(E) The twentieth century A.D.

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	16	9	10	5	3
A	6	7	9	7	6
B	9	8	1	5	
C	8	6	5	3	
D	5	3	5	2	3
*E	27	41	44	52	62
Total	71	74	74	74	74

Per cent of total group of 370 students answering correctly 62%
Correlation between success on this question and total score on humanities section of test ..... .40

(Discrepancies in total numbers in each fifth are caused by drop-out of students not completing the test.)

*Correct answer

This question on art appeared in a test of general background given to college seniors and graduate students. To answer the question, the student must apply his knowledge of the characteristics of various periods in the history of sculpture in order to place the statue within its proper period.

The statue can be identified immediately as a product of western civilization because of its subject matter and design. The sculpture most commonly associated with western art in the fifth century B.C. is that done in Greece. During that period Greek sculpture showed its characteristic idealization of the human figure. If the student contrasts a well-known statue from that period, such as Myron's "Discobolus" ("The Discus Thrower") with the statue in the picture, he can easily discover that the treatment accorded the two figures is not the same. The figure of the woman pictured is not idealized, and although the frontal pose is found in Greek sculpture, a turned head on the figure is not. Therefore, choice (A) can hardly be correct.

In some countries sculptors of the fourteenth century A.D. began to create statues for use in palaces, chapels, and tombs of private individuals, but their work retained the qualities of the representations of saints found in the Gothic cathedrals of Europe. Statues of men and women associated with Christian history were also used as architectural decorations. In the fourteenth century probably only Eve would have been carved as a nude female figure, and she would certainly have been portrayed in a more modest pose and with a fig leaf. Although, like the figure of the woman pictured, the figures of this period of Christian art are not idealized, the statue of the woman displays none of the other characteristics of medieval art, and choice (B) can be eliminated.

The fifteenth century saw the beginnings of the Italian Renaissance, and by the sixteenth century sculptors were again presenting idealizations of the human figure, most often in a Christian rather than classical context. The statue pictured above shows none of the heroic character that is found in the works of Michelangelo and none of the exaggerated stylization that appeared in the works of his followers. Thus, choice (C) can be regarded as inaccurate.

Early in the eighteenth century sculptors frequently displayed a self-conscious coy femininity in their female nudes, but by the end of that century sculptors such as Canova had adopted the classical ideals and imitated classical models. Consequently, choice (D) is not a valid one, for the statue in the photograph does not resemble an eighteenth-century interpretation of a Greek goddess or nymph. It does, however, realistically present a human figure without any attempt at idealization. The realism here is thoroughly modern. The use the sculptor makes of texture points to the modern period, for it was only late in the nineteenth century that sculpture (such as Renoir's "The Washerwoman") used nonrepresentational texture as an integral part of the design and aesthetic effect of the work of art. Minor details in the statue such as the shape of the nose, the half-closed eyes (instead of open sockets), and the line of the hair indicate that the work is not classical in origin, and at the same time they point to the twentieth century school of art. The statue is actually "Junge Frau," the work of the modern German sculptor, Georg Kolbe.

The statistics show that the question was an easy one for the particular group that took the test. The correct answer was chosen by 62 per cent of the students. However, more of the students in the highest group than in the lowest group selected the correct response, and the incorrect answers were generally more appealing to the poorer students than they were to the better students.

Question 6

In the following question you are given a complete sentence to be rephrased according to the directions which follow it. You should rephrase the sentence mentally to save time, although you may make notes in your test book if you wish.

Below the sentence and its directions are listed words or phrases that may occur in your revised sentence. When you have thought out a good sentence, find in the choices A to E the word or entire phrase that is included in your revised sentence. The word or phrase you choose should be the most accurate and most nearly complete of all the choices given.

Although the directions may require you to change the relationship between parts of the sentence or to make slight changes in meaning in other ways, make only those changes that the directions require; that is, keep the meaning the same, or as nearly as the directions permit. If you think that more than one good sentence can be made according to the directions, select the sentence that is most exact, effective, and natural in phrasing and construction.

Sentence: **John, shy as he was of girls, still managed to marry one of the most desirable of them.**

Directions: **Substitute John's shyness for John, shy.**

Your rewritten sentence will contain which of the following?

(A) him being married to

(B) himself married to

(C) him from marrying

(D) was himself married to

(E) him to have married

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	6	1	2		
A	4	2	5	1	3
B	9	6	2		
*C	9	21	33	44	45
D	11	8	3	2	
E	11	12	5	3	2
Total	50	50	50	50	50

Per cent of total group of 250 students answering correctly 61%
Correlation between success on this question and total score on test 66

*Correct answer

In developing this type of question the committee of examiners reasoned that most good writers reconstruct sentences to change emphasis, to improve style, to avoid ambiguity, or to eliminate verbosity, and that the good student should be able to make such changes without involving himself in structural faults or grammatical errors. This type of question is designed, therefore, to assess the student's mastery of variety in sentence structure, his ability to make a change within a sentence so that it says what he intends to say more smoothly, concisely, and effectively than the original version may. Furthermore, the committee decided that to test this ability the question should require the student first to construct his new sentence mentally and then to compare his answer with a number of possible answers presented to him. As choices for answers they chose the kinds of expressions which students include in their own writing when they attempt to solve problems of sentence variety and become enmeshed in grammatical incongruities or verbal obscurities.

In order to select choice (A), a sentence like John's shyness with girls did not stop him being married to the most desirable of them would have to be used. To make this sentence correct, formal written English demands that the word being be preceded by the possessive pronoun his. Choice (B) presents a sentence similar to Despite John's shyness with girls, he managed to get himself married to one of the most desirable of them. This sentence is wordy and inappropriate in its tone (for formal English). Choice (C), however, yields a sentence on the order of John's shyness with girls did not prevent him from marrying one of the most desirable of them. This retains the meaning of the original sentence and contains no errors in grammar; it is the correct answer. The fourth choice might lead to John's shyness with girls did not keep him single; he was himself married to one of the most desirable of them. This sentence changes the meaning of the original sentence, and it is, at the same time, ambiguous in its own meaning. John's shyness with girls was not a reason for him to have married the most desirable of them—this sentence, an attempt to use the fifth choice, results in a complete change of meaning. It is therefore unacceptable, even though it is grammatically correct.

This question was answered correctly by 61 per cent of the college-bound high school students who attempted it. Most of the students in the two highest fifths answered it correctly, while a majority of students in the two lowest groups answered the question incorrectly.

Question 7

The sentence below has blank spaces, each blank indicating that a word has been omitted. Beneath the sentence are five lettered sets of words. You are to choose the one set of words which, when inserted in the sentence, best fits in with the meaning of the sentence as a whole.

From the first the islanders, despite an outward—, did what they could to—the ruthless occupying power.

(A) harmony .. assist
(B) enmity .. embarrass
(C) rebellion .. foil
(D) resistance .. destroy
(E) acquiescence .. thwart

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit		2	4	1	
A				1	
B	4	1			
C	10	7	2		2
D	37	28	16	8	3
*E	9	22	38	50	55
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 58%
Correlation between success on this question and total score on test 69

*Correct answer

This type of question is one of several commonly used in tests in which a measure of verbal aptitude is sought. Specifically, this type is designed to measure one aspect of reading comprehension: the student's ability to recognize the consistency in logic and in style that is required of the elements in a sentence. If the student understands the implications of a sentence, he should be able to select the answer that best fulfills the meaning of the sentence. The sentences deal with a wide variety of topics that the student is likely to have encountered in general reading.

Examination of the question should indicate that the answer involves two words which are, in a sense, opposite in meaning, since the word "despite" carries with it the implication that the islanders acted in one fashion, while presenting a somewhat different impression to the "ruthless occupying power." With this relationship in mind, (A), (B), (C), and (D) can be eliminated since all of those answers fail to give the sense of contrast that is required. If outward harmony existed, assisting the occupying power would probably contribute to this state of harmony and strengthen it. If enmity existed, embarrassing the occupying power would be one method of expressing this feeling. Should rebellion exist, then the islanders would be doing what they could to foil the occupying power; these terms do not imply opposition. The same logic holds for "resistance . . destroy." The correct answer, the only one implying two opposed actions, is (E).

In this analysis, 58 per cent of a group of college-bound secondary school students answered the question correctly, and it discriminated extremely well between the good students and the poor students. Performance on this question

correlated .69 with performance on the test as a whole. The analysis shows that whereas nearly all of the top one-fifth answered correctly, only 15 per cent of the lowest one-fifth did so.

Questions 8-9

The passage below is followed by questions based on its content. After reading the passage, choose the best answer to each question. Answer all questions following the passage on the basis of what is stated or implied in the passage.

Mill's estimate of Harriet's gifts may have been extravagant. Their contemporaries, few of whom knew her well, were inclined to question it. As Goldwin Smith waspishly remarked, "Mill's hallucination as to his wife's gifts deprived him of all authority wherever that came in." But there can no longer be substantial doubt of the range of her influence. Mr. Packe corroborates the conclusion reached by Professor Hayek in his introduction to the Mill-Taylor correspondence—that Mill was only stating sober fact when he credited Mrs. Taylor (who later became Mrs. Mill) with a large share in the work published under his name. Nor was it a matter of feminine flattery or, as Carlyle imagined, of "those great dark eyes, that were flashing unutterable things while he was discoursin' the unutterable." Mr. Packe insists that, except for the Logic, the principles underlying the more important works of John Stuart Mill were defined although not actually composed by Harriet Taylor, and that every major work after the Political Economy was drafted or planned during their first few years of married life. This is not necessarily to accept Mill's view of his wife's mental endowments. She was obviously a woman of lively mind and exceptional charm to whom Mill responded intellectually as well as emotionally. But we may still suspect that her ascendancy over him was, at bottom, one of those mysterious functions of personality rather than of intellect.

Question 8

According to Mr. Packe, Harriet's part in Mill's important works was

(A) mainly restricted to small details

(B) in the planning rather than in the execution of them

(C) in the comfort of her personality and charm

(D) only evident in the Logic

(E) helping with the more difficult sections

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	4	11	4	2	1
A	1	3	1	1	
*B	19	34	51	46	64
C	14	8	5	11	4
D	22	6	3	2	
E	14	12	10	12	5
Total	74	74	74	74	74

Per cent of total group of 370 students answering correctly 58%
Correlation between success on this question and total score on test . . . 51

*Correct answer

Question 9

This passage shows that Goldwin Smith's opinion of Mill was that Mill

(A) was naturally the person who would know best about his own wife's intellectual powers

(B) had no longer any authority over his wife

(C) was able by some sixth sense to see that his wife had outstanding intellectual ability

(D) was very seriously mistaken about his wife's intellectual prowess

(E) was not the real "architect" of the writings published in his name

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	2	7	5	1	1
A	12	3	1	3	
B	27	29	30	22	23
C	6	5	5	2	
*D	8	12	16	36	42
E	19	18	17	10	8
Total	74	74	74	74	74

Per cent of total group of 370 students answering correctly 31%
Correlation between success on this question and total score on test . . . 48

*Correct answer

Question 8 is fairly easy, and 58 per cent of a group of college-bound high school seniors answered it correctly. The question clearly asks for a restatement of Mr. Packe's opinion as to the nature of Harriet Mill's contribution to her husband's work. In lines 6-13, we learn that Packe corroborates Professor Hayek and John Stuart Mill himself in the view that Mrs. Mill had a large share in the work published by Mill. In lines 16-22, we learn that Packe considers this contribution to consist of planning the work in outline. Accordingly, only (B) correctly states his opinion. Choice (A) is wrong because it suggests only a very small role for Mrs. Mill. Choice (C) fails to suggest any real intellectual contribution, which is counter to Packe's view that the prin-

ciples underlying the works were defined by Mrs. Mill. Choice (D), which was the most popular choice of the least able group, states that Mrs. Mill's contribution is only evident in the *Logic*, whereas the passage reports that Packe specifically excepted the *Logic* from his remarks. Finally, choice (E), while suggesting an important role for Mrs. Mill, goes far beyond the passage and describes her work as helping when difficulty arose. The passage nowhere suggests that she had such a role. The question correlates well with test score, as is evident from the increasing proportion of candidates who answer it successfully as the ability level increases.

Question 9 proved to be more difficult for the group (the two analyses are derived from the same candidates). It is much more specific, being based on the sentence in lines 3-6 which quotes the remark of Goldwin Smith. Smith, of course, is taking violent exception to Mill's opinion of his wife's gifts. He refers to this opinion as a "hallucination," and says that it deprives Mill of all "authority," i.e., all right to speak with authority, on that topic.

Clearly, then, the correct restatement of Smith's opinion of Mill is (D), which says that Mill was mistaken about his wife. Choice (A) is directly opposite in meaning, and is clearly wrong. Choice (B) is wrong, for it misinterprets "deprived him of all authority wherever that came in." It is the most common mistake, and many quite able candidates made it, but it is clearly wrong. Choice (C) is wrong in the same sense in which (A) is wrong: it is too positive a thing to say. Finally, choice (E) is wrong because it is completely beyond the scope of Smith's remark, which had nothing to do with the question of who planned Mill's work.

As in the first example, the question shows a definite tendency for the proportion of successful candidates to increase as the ability level rises.

Questions 10-11

Questions 10-11 are based on the following German reading passage:

Als in der ersten Nachkriegszeit eine Flut von billiger Schundliteratur unser Land überschwemmte—(in einem einzigen Zuricher Schulhaus fand man bei einer Durchsuchung tausend solcher Heftchen)—bildete sich zur Bekämpfung dieser Gafahr eine private Arbeitsgemeinschaft. Da für Verbote keine gesetzliche Grundlage vorhanden war, beschritt man den Weg, der sich noch immer als der beste bewährt hat: man suchte das Schlechte durch das Gute aus dem Felde zu schlagen. Es sollte moralisch und literarisch einwandfreie, spannende, lehrreiche, zu eigenem Tun anregende Jugendlektüre zu bescheidenem Preis geschaffen werden. Da kein schweizerischer Verlag dafür zu haben war, grundeten die Initianten im Jahre 1931 einen eigenen Verlag. Pro Juventute übernahm die Geschäftsstelle, und die schweizerische Lehrerschaft setzte sich bis ins letzte Bergdorf mit unermüdlichem Eifer für das Werk ein, das heute in der Erziehung der Schweizer Jugend einen bedeutenden Platz einnimmt.

Translation:

When our country was swamped by a flood of trashy literature shortly after the war (a search of a single schoolhouse in Zurich turned up a thousand copies of such publications), a private cooperative was formed to combat this menace. Since there was no legal basis for suppressing these writings, the problem was approached in a way that has always proven to be the most successful—that of using the good to rout the bad. Reading matter of unquestioned moral and literary value was to be made available to young people at modest prices, reading matter both suspenseful and instructive, which would stimulate young people to independent action. Being unable to interest a Swiss publishing house in the project, the promoters founded their own publishing house. Pro Juventute assumed management, while the entire teaching profession—down to the last mountain village—committed itself wholeheartedly to this undertaking, which now occupies a central position in the education of Switzerland's young people.

Question 10

Pro Juventute wurde gegründet, um

(A) der schweizerischen Lehrerschaft ein Einkommen zu verschaffen

(B) die Gesetze des Landes zu umgehen

(C) ausländische Bücher vom Markte zu verdrängen

(D) Überschwemmungen des Landes zu verhüten

(E) wertvolle und preiswerte Jugendliteratur zu veröffentlichen

Translation:

Pro Juventute was founded in order to

(A) provide income for the Swiss teaching profession

(B) circumvent the laws of the land

(C) squeeze foreign books from the market

(D) prevent floods

(E) publish worthwhile and moderately priced reading material for young people

Statistical Analysis

	Students Classified by Total Test Score					
RESPONSES	LOWEST GROUP	NEXT LOWEST GROUP	THIRD LOWEST GROUP	THIRD HIGHEST GROUP	NEXT HIGHEST GROUP	HIGHEST GROUP
Omit	28	35	23	16	4	
A		5	6	2		
B		4	6	3		
C		6		2		
D	3	6	3		2	
*E	1	15	24	13	24	65
Total	32	71	62	36	30	65

Per cent of total group of 296 students answering correctly 48%
Correlation between success on this question and total score on test 80

(The statistical analysis presented for this question differs somewhat from those previously presented. In this type of analysis the total group is divided into six sub-groups so that if the scores are distributed according to the normal curve the numbers in each sub-group will be equal. In this particular case, however, the scores were not normally distributed; thus the sub-groups are not equal in size.)

*Correct answer

Question 11

Man konnte die minderwertige Literatur nicht verbieten, weil

(A) sie privates Eigentum war

(B) es keine bessere Jugendlektüre gab

(C) sie nicht gesetzwidrig war

(D) sie billig und weit verbreitet war

(E) sie dem Geschmack der Nachkriegszeit entsprach

Translation:

The trashy literature could not be suppressed because

(A) it was private property

(B) no better reading material for young people was available

(C) it was not illegal

(D) it was cheap and widely read

(E) it satisfied public taste in the postwar period

Statistical Analysis

	Students Classified by Total Test Score					
RESPONSES	LOWEST GROUP	NEXT LOWEST GROUP	THIRD LOWEST GROUP	THIRD HIGHEST GROUP	NEXT HIGHEST GROUP	HIGHEST GROUP
Omit	29	38	25	23	5	1
A	1	6	9	1		1
B	1	7	10	3	7	
*C	1	6	7	6	10	61
D		5	6	2	4	2
E		9	5	1	3	
Total	32	71	62	36	29	65

Per cent of total group of 295 students answering correctly 31%
Correlation between success on this question and total score on test 92

(As in the case of the analysis for the preceding question, the statistical analysis presented for this question is based on a somewhat different procedure.)

*Correct answer

This passage was included in a German test for college graduates applying for government service. A number of multiple-choice questions were based on this passage in order to ascertain whether the applicant comprehends with accuracy the factual information conveyed by this piece of straightforward newspaper reporting and whether he also understands the implications of important details as well as the author's purpose in writing this news story. The ability to understand such a highly idiomatic expository text was considered essential for the group for which the test was intended.

In order to succeed on this kind of a test, the candidate needs a fair command of basic German vocabulary and of word structure and sentence structure, as well as some experience in reading contemporary German prose. He is not expected to know every word in the text, but rather to be able to infer the meaning of unfamiliar expressions from the context.

From the group of comprehension questions based on this passage, two have been selected for detailed discussion in order to demonstrate the techniques used in good foreign-language testing. Basically, the questions are designed to give the examinee an opportunity to show what he can do with the knowledge and the skills he has acquired, thus testing this knowledge indirectly through its application to problems posed to him.

Question 10 tests total comprehension of the passage. The question asked is, "For what purpose was Pro Juventute founded?" All of the five answer choices contain some elements derived from the text, but only the one correct answer is in complete agreement with the information offered in the passage. Those who understood certain words or phrases but failed to comprehend the main point were led to choose one of the wrong answers. The statistical analysis shows that choice (A) attracted 13 candidates, most of whom performed rather poorly on the test as a whole. They vaguely understood that the teachers engaged in a business activity but inferred erroneously the teachers' motives for doing so. Choice (B) attracted a similarly poor group of 13 candidates who based their responses on their recognition that the matter of legality was discussed (lines 5-9), but who failed to understand its meaning in the context of the passage. Choice (C) is based on a knowledge of only a few scattered words or phrases: "books," "market," "no Swiss publishing house available." It attracted a few examinees who had only a piecemeal knowledge of vocabulary and a minimum of ability to comprehend meaning. A small group of examinees was led by the superficial connection through the word "flood" in the first line of the passage to pick (D). The correct choice (E) was selected by only one examinee of the lowest group and *all* 65 examinees in the top group. The percentage of those choosing (E) increases markedly with the ability of the sub-group. Hence the excellent correlation coefficient of .80.

Question 11 deals with an important detail. Why was it impossible to suppress the trashy literature? Here again, each answer choice contains an element of information from the passage; but as a response to the question implied, only choice (C) makes sense. This question, with its unusual correlation coefficient of .92, represents about the maximum discriminating power that can be attained with a single question. It is a strikingly effective measure of whether an individual can operate in German as effectively as in his native language.

Question 12

Directions: Maintenant, vous allez entendre une conversation entre deux personnes. Attendez la deuxième réplique et ensuite choisissez la réponse qui convient le mieux.

"Henriette, passe-moi cette petite robe légère qui se trouve dans mon armoire à glace."

"Attends un moment; j'ai la bouche pleine de pâte dentifrice. J'aurai fini ma toilette dans un instant."

(A) Ne te presse pas. Je la chercherai moi-meme.

(B) Je l'ai trouvée tout à l'heure.

(C) Quand tu auras fini de te peigner.

(D) Qui, je l'ai repassée hier soir.

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	3	1		1	
*A	21	29	45	57	59
B	14	4	1	2	
C	19	23	13		1
D	3	3	1		
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 70%
Correlation between success on this question and total score on test . . 65

*Correct answer

The testing of a student's ability to understand a foreign language when spoken is a relatively recent development. Students listen to the recorded voices of native speakers in statements, short conversations, and short narrations. Suggested answers to questions based on the spoken material are printed in the students' test books. In listening comprehension tests the foreign language is used throughout, not only in the test questions but also in the directions to students. The sample question given here was part of a test designed for teachers of French in secondary schools. Paraphrased in English it reads:

Directions: Now you will hear a conversation between two people. Wait for the reply of the second speaker and then select the reply which is most appropriate.

Recorded conversation:

"Henrietta, hand me the little summer dress which is in my closet."

"Wait a moment; I have a mouth full of toothpaste. I shall be through in a minute."

Answer choices:

(A) Don't hurry. I shall get it myself.

(B) I found it just a little while ago.

(C) When you are through doing your hair.

(D) Yes, I ironed it last night.

To select the right answer to this question, the student has to apply his knowledge of French grammar, vocabulary, and idiomatic expression, and his understanding of the French sound system. The students who have grasped the meaning of what they have heard will select (A) as the only choice that makes sense in the context. The literal use of the French word "trouver" (to find) in choice (B) misleads some who have not understood that the idiomatic expression "se trouver" in the conversation means simply "to be in a certain place." Choice (C) misleads those who have understood that the second woman will soon be finished with whatever she is doing but have not understood the nature of her activity. Choice (D) was selected by some students who mistook "passer" (to hand) in the conversation for "repasser" (to iron). This question was answered correctly by 70 per cent of those who attempted it. All but one of the top fifth selected the right answer, whereas only 35 per cent of the lowest fifth did so.

Question 13

The concept of the plasma membrane as a simple sieve-like structure is inadequate to explain the

(A) passage of gases involved in respiration into and out of the cell

(B) passage of simple organic molecules such as glucose into the cell

(C) failure of protein molecules to pass through the membrane

(D) inability of the cell to use starch without prior digestion

(E) ability of the cell to admit selectively some inorganic ions while excluding others

Statistical Analysis

RESPONSES	Students Classified by Total Test Score				
	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	4	4	2		
A	3		1		
B	2	4		1	
C	7	3	1	3	1
D	1	1	1	1	1
*E	24	29	36	36	39
Total	41	41	41	41	41

Per cent of total group of 205 students answering correctly 80%
Correlation between success on this question and total score on test . . 48

*Correct answer

This question was asked of high school students who were completing a college-level course in general biology. The question was a relatively easy one for the group, as evidenced by the fact that 80 per cent of the students selected the correct answer. The ability to handle this question successfully involves the knowledge that the *living* plasma membrane has properties in addition to those served by a piece of cellophane or similar material used in classical laboratory demonstrations of osmosis. A membrane such as cellophane can indeed be thought of as a porous structure. Only those molecules that are smaller than the largest of the pores would be able to pass through, accounting fully for the fact that both gases and glucose, whose molecules are relatively small, do pass through. The concept of the plasma membrane as a sieve also accounts adequately for the fact that protein molecules cannot pass through, since protein molecules are among the largest known. Choice (D) is a poor choice since it is well known that starch molecules are very large; as one might expect, only a few of the students in the sample selected this choice. Experiments have shown that certain ions of inorganic salts, acids, and bases, even though these ions are small, penetrate some plasma membranes very slowly, but pass readily through other plasma membranes having pores of the same size. If only the "sieve" nature of the membrane accounted for the selective passage of materials through it, one would expect these ions to pass through all membranes rapidly. To account for the actual phenomenon, then, one must have recourse to more elaborate interpretations, notably that electrical properties of the living membrane selectively attract and repel certain ions and that the chemical nature of the membrane constituents selectively allows passage of materials that are

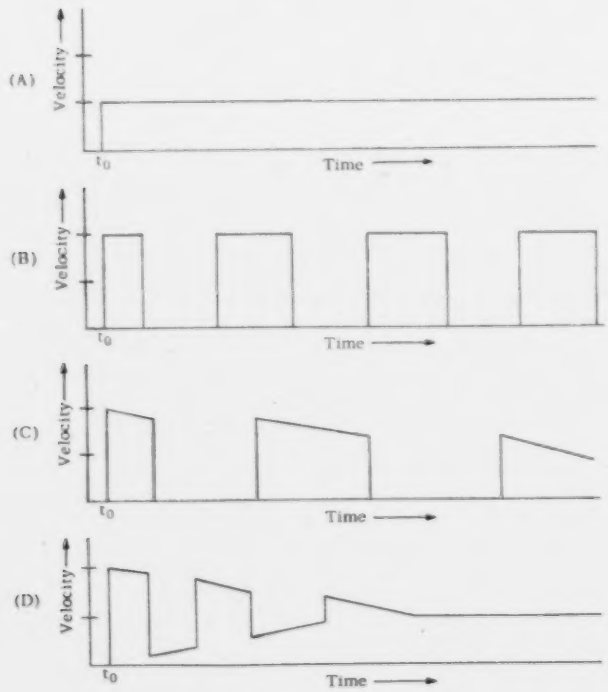
soluble in these constituents. The statistical analysis shows that better students selected (E) as the only phenomenon not explainable by the conception of the plasma membrane as a simple sieve.

Questions 14-15-16

Questions 14-16 relate to the following information and graphs; they consist of five lettered answer choices followed by a list of numbered questions. For each question select the one lettered answer which is most closely related to it. An answer may be used once, more than once, or not at all.

A standard impulse is given at time t_0 to each of the several closed boxes having identical masses and identical appearance. The impulse starts the box into motion on a horizontal plane. There may or may not be friction (referred to as "external friction") between any box and the plane. It is known that each box contains a disk whose mass is equal to that of the box and which is resting on the bottom of the inside of the box. There may or may not be friction (referred to as "internal friction") between any disk and the bottom of its box. The disks may rebound in a perfectly elastic way from the walls of the box ("elastic walls"), or they may stick to the walls on colliding ("sticky walls").

The graphs below show the velocities of various boxes as a function of time.



(E) None of the above

Question 14

Which of the graphs could represent the velocity-time relationship of a box that has no external friction and whose disk is fastened solidly to it?

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	1	1			
*A	55	69	72	73	73
B	11	1			
C	2	1		1	
D	3	1	1		
E	2	1	1		1
Total	74	74	74	74	74

Per cent of total group of 370 students answering correctly 92%
Correlation between success on this question and total score on test . .56

*Correct answer

This set of questions appears in one of the achievement tests developed to accompany the high school physics course of the Physical Science Study Committee.

Question 14 is the first of several questions based on the situation described. The series of questions requires a knowledge of certain aspects of kinematics and dynamics: an understanding of Newton's laws of motion and the concepts of velocity, acceleration, and momentum. The student must apply this knowledge to an unfamiliar problem and answer the questions in terms of graphs, a frequently used tool of the physical scientist.

The situation described is an idealized one since the complications of friction are occasionally ignored. The study of phenomena as they would occur under certain idealized conditions is a major way to attack complex scientific problems.

Question 14 pertains to a box with no external friction and with a disk fastened solidly to the box. When a standard impulse is given to the box with its attached disk at time t_0, the box would be set in motion. Since no friction or other force acts on the box after the initial impulse, the box would neither accelerate nor decelerate but would continue in motion at constant velocity. Choice (A) illustrates this motion.

The question was extremely easy; 342 of the 370 students (92 percent) who attempted it answered it correctly. Nonetheless, the correlation between performance on this question and the test as a whole is high (.56).

The most discriminating incorrect answer choice was (B). Several students in the lowest fifth on the test seemed to recognize that in a frictionless situation, like that of this question, the velocity would not change with time, but they failed to recognize the simplicity of this situation and chose (B), where the velocity pattern is constant with time but where the velocity regularly drops to zero for short intervals.

Question 15

Which of the graphs could represent the velocity-time relationship of the same box with external friction?

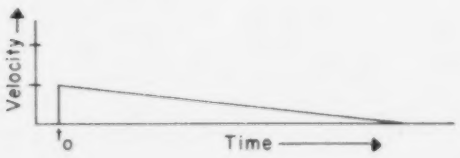
Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	2	2		1	
A	9	5	2	4	4
B	22	10	9	4	1
C	28	27	29	15	3
D	7	7	1	1	
*E	6	23	33	49	66
Total	74	74	74	74	74

Per cent of total group of 370 students answering correctly 48%
Correlation between success on this question and total score on test . . . 70

*Correct answer

One condition is now added to the situation described in the preceding question. The student is asked to choose the graph which could represent the motion of the box *with* external friction. The force of friction would act to decelerate the box and the attached disk at a constant rate. Therefore, the velocity-time relationship of the box would be:



Since none of the graphs given represents this motion, the answer is (E), None of the above.

This question was considerably more difficult than question 14; only 48 per cent of those who attempted it answered it correctly. One must determine what the velocity-time relationship would be and then recognize that the correct graphical relationship is not offered as a choice.

The correlation between performance on the total test and performance on this question is very high, .70. (B) and (C) were the most common choices of students with low scores on the test. These students apparently failed to realize that, as long as the disk is fastened to the box, the situation is as simple as if the disk were not present. The larger number of them chose (C), where the maximum velocity does drop with time, but both (B) and (C) have the weakness that they are appropriate only for more complex situations.

Question 16

A box with elastic walls and no internal or external friction has its disk initially resting in the middle of the box. Which of the graphs could represent the velocity-time relationship of this combination?

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	2	2			
A	17	10	5	4	
B	18	28	44	53	69
C	15	10	6	4	1
D	16	17	16	11	2
E	6	7	3	2	2
Total	74	74	74	74	74

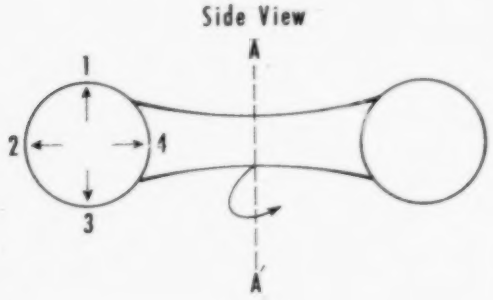
Per cent of total group of 370 students answering correctly 57%
Correlation between success on this question and total score on test . . . 61

*Correct answer

When the standard impulse is given to the box described in this passage, the box moves until the disk inside, free to slip along the bottom since there is no friction, hits the trailing end of the box. At this point, there is an elastic collision between the box and the disk, which have equal masses. All the momentum of the box is transferred to the disk, and as a result, the box stops and the disk moves across the bottom of the box. The disk soon hits the forward end of the box in a second elastic collision. The momentum is once more transferred to the box, the disk stops, and the box once again moves at the same velocity as before. The velocity of the box is undiminished since there are no frictional forces acting to slow the motions. Choice (B) represents this situation. This question was answered correctly by 57 per cent of the examinees. The correlation between performance on this question and performance on the total test in which it was used was high, .61. Only five of the students with the highest fifth of the test scores failed to analyze the situation correctly and to select the correct answer. Students with the lowest scores apparently did not understand the situation. Although very few of them omitted the question, their responses are distributed among choices (A), (B), (C), and (D) as if they were answering at random. Few of the students with the lowest scores chose the "None of the above" response, indicating their belief that the correct graph was presented, but they were, in general, not able to decide which one it was.

Question 17

One method of obtaining "artificial gravity" in a space station is to have the station rotating about axis AA as it revolves around Earth.



The inhabitants of the space station would call which direction "down"?

(A) Direction 1

(B) Direction 2

(C) Direction 3

(D) Direction 4

(E) Any one of the four, depending on speed of rotation

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	3		1		
A		1		1	
*B	12	17	19	29	28
C	8	4	6	1	
D	7	9	6	3	6
E	4	3	2		
Total	34	34	34	34	34

Per cent of total group of 170 students answering correctly 62%
Correlation between success on this question and total score on test . . . 44

*Correct answer

This was part of a set of questions administered to high school students who were completing a year of high school physics. The question illustrates the kind of response that can be expected of well-trained students of high school age when they are presented with a relatively novel situation which is based on fundamental concepts from the field of mechanics.

This question requires that the student consider the nature of a possible mechanism for providing a "down" direction in a space station to simulate the gravitational "down" so important in our normal activities on Earth. Choice (C) is the direction normally considered down in diagrams. Although this direction is not significant in the space station, a sizeable number of the poorer physics students chose it. Other students assumed that the "down" direction would be that one toward the center of rotation of the station, choice (D). However, objects free to move in the space station behave as do particles in a centrifuge and "fall" to the outer edge. This direction, (B), then is the "down" direction in the rotating station. The other choices, (A), "up" as it is usually represented in diagrams, and (E), a direction which depends on the speed of rotation, were not selected by many students.

Question 18

Intensity of color may be used as a measure of concentration. The permanganate ion MnO_4^- is intensely purple. If the following salts were separately dissolved in water to make 0.001-molar solutions, the solution having the most intense color would be

(A) $KMnO_4$

(B) $NaMnO_4$

(C) $NaMnO_4 \cdot 3H_2O$

(D) $Ca(MnO_4)_2$

(E) $AgMnO_4$

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	26	18	15	7	8
A	13	16	14	16	14
B	6	5	8	4	3
C	2	1	5	6	3
*D	8	13	11	22	31
E	5	7	7	5	1
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 28%
Correlation between success on this question and total score on test . . . 38

*Correct answer

This question, which was asked of high school chemistry students, involves a rather simple concept with which all students at this level would be expected to be familiar. The concept might be stated as follows: Even though the number of moles in a liter of solution of each of several salts is the same, the concentrations of the ions of these salts need not be the same. A mole of $Ca(MnO_4)_2$ provides two moles of MnO_4^- ions while a mole of each of the other salts listed provides only one mole of MnO_4^- ions. As a result, when the numbers of moles of the salts present in solution are the same, a solution of $Ca(MnO_4)_2$ contains more moles of MnO_4^- ions than any of the other solutions, and is therefore more intensely colored.

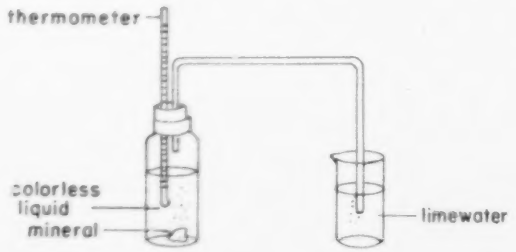
The setting of the question is somewhat unusual, and the question is, therefore, useful in discriminating between students who are able to see the application of a familiar concept in an unusual situation and those who are not. About 30 per cent of the students selected the correct response (D). Of the other compounds listed, the most familiar one is $KMnO_4$, choice (A). A number of students apparently were unable to see the application of any concept in this situation and so selected the substance whose formula was most familiar, assuming that a familiar compound is most likely to be right. Nearly half of the lowest fifth of

the students omitted the question, apparently because they did not realize that a familiar concept would lead to the answer. Students who chose (B), (C), and (E) demonstrated that they did not understand the meaning of "mole": (B) and (C) list the same salt, first in anhydrous form and then hydrated; (E) lists a salt of a heavy metal with a resulting high molecular weight. These differences among the salts do not affect the number of MnO_4^- ions introduced into the solution by a mole of each salt and thus choices (B), (C), and (E) are incorrect.

Question 19

This question is based on the following situation:

A piece of mineral is placed in a bottle half-filled with a colorless liquid. A two-holed rubber stopper is then placed in the bottle. The system is then sealed by inserting a thermometer and connecting a glass tube to the stoppered bottle and a beaker of limewater as shown in the accompanying diagram:



The following series of observations is recorded:

I. Observations during the first few minutes:
1. **Bubbles of a colorless gas rise to the top of the stoppered bottle from the mineral.**
2. **Bubbles of colorless gas begin to come out of the glass tube and rise to the surface of the limewater.**
3. **The limewater remains colorless throughout this period of time.**
4. **The thermometer reads 20°C.**

II. Observations at the end of thirty minutes:
1. **Bubbles of colorless gas continue to rise in the stoppered bottle.**
2. **The piece of mineral has become noticeably smaller.**
3. **There is no apparent change in the level of the colorless liquid in the bottle.**
4. **The colorless liquid in the bottle remains colorless.**
5. **The thermometer reads 24°C.**
6. **The limewater is cloudy.**

Which one of the following is the best explanation for the appearance of gas bubbles at the end of the tube in the beaker of limewater?

(A) The pressure exerted by the colorless liquid is greater than that exerted by the limewater.

(B) The bubbles coming from the mineral cause an increased gas pressure in the stoppered bottle.

(C) The temperature increase at the end of thirty minutes causes an expansion of gas in the stoppered bottle.

(D) The decrease in the size of the piece of mineral causes reduced pressure in the stoppered bottle.

(E) The glass tube serves as a siphon for the flow of gas from the bottle to the beaker.

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit		2			
A	10	4	3	5	1
*B	13	28	34	43	52
C	11	13	4	5	1
D	5		1		
E	21	13	18	7	6
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 57%
Correlation between success on this question and total score on test ... 55

*Correct answer

This question is taken from a test designed to be used with a new curriculum in high school chemistry. The question is only one of a series based on the experimental situation described. Questions in the series are grouped in sequence relating to the situation in order to permit the student to think intensively in one setting for an extended period of time. The student is asked to deal with a realistic laboratory situation—one he has not yet encountered in the course at the time the test was given—and to employ scientific problem-solving ability in using the data given to answer the questions.

Choice (A) is both vague and irrelevant. It is unspecific about where the pressure is exerted and has nothing to do with the cause of the bubbles.

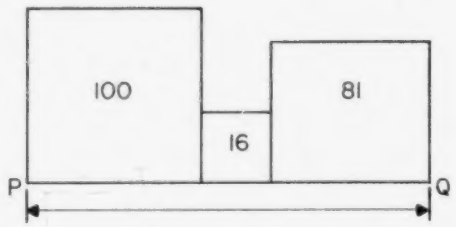
Choice (C) sounds plausible. In itself, the statement is not incorrect, since the temperature increase will cause an increase in gas pressure in the stoppered bottle. However, students who are in command of the subject will realize that the increase in gas pressure due to the rise in temperature is insignificant compared with that caused by the bubbles coming from the mineral.

Choice (D) embodies an incorrect statement. Some students may not realize this, but even if the statement were correct, it would offer an incorrect explanation for the appearance of the bubbles in the beaker.

Most of the good students rejected choice (E), but many less able students chose it, probably because the arrangement looks somewhat like a siphon and they had heard of the use of siphons in transferring fluids.

Students who understand the forces at work in the situation described will know that bubbles would appear at the end of the tube in the beaker when the pressure exerted by the gas from the stoppered bottle exceeded the pressure exerted by the limewater at the end of the tube. They will also realize that the limewater pressure would remain essentially constant. Choice (B), therefore, sets forth the condition that would account best for the appearance of gas bubbles at the end of the tube in the beaker.

Question 20



In the figure above, three squares with areas of 100, 16, and 81 lie side by side as shown. By how much must the area of the middle square be reduced in order that the total length PQ of the resulting three squares be 21?

(A) $\sqrt{2}$ (B) 2 (C) 4 (D) 8 (E) 12

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	30	12	12	4	1
A	8	9	8	2	1
B	11	18	8	10	3
C	4	3	4	3	4
D	2	3	3	1	
*E	5	15	25	40	51
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 45%
Correlation between success on this question and total score on test . . . 66

*Correct answer

This question is similar to those included in tests of mathematical aptitude for high school students. One of the problems surrounding such aptitude questions is the need to avoid familiar textbook material while at the same time restricting the content level to that which was mastered earlier in the junior high school—geometry and simple algebra. If this need is met, then the focus of the aptitude test is where it should be, namely on ingenuity in the solving of novel problems, rather than on the recall of memorized procedures. This ingenuity, or what some like to call "mathematical maturity," probably improves with taking good mathematics courses throughout high school but is not much affected by short-range cramming.

This question involves procedures which the student is extremely unlikely to have encountered in high school textbooks; the content knowledge required is certainly not beyond what is taught in the junior high school. The solution proceeds as follows:

The lengths of the three sides of the squares are 10, 4, and 9, reading from left to right. Therefore, initially PQ equals 23. If the final length of PQ is to be 21, the amount of reduction of the side of the middle square must be 2. The middle square must consequently be reduced from an area of 16 to an area of 4; that is, it must be reduced by 12 square units.

Of the 59 college-bound high school students in this sample who omitted the question, 54 were in the lowest three-fifths. The 50 who chose (B) were students who started toward a correct solution but who stopped with the length of the new side of the middle square. The 18 choosing (C) seem to have proceeded as far in their solution as finding the new area of the middle square and, on the average, this was the most able of the groups missing the question. Choices (A) and (D) result from taking wrong directions in the solution of the problem and, as expected, these were selected by the least able groups.

Question 21

What are the roots of $x^2 - \left((a + \frac{1}{a}\right)x + 1 = 0$?

(A) $a, \frac{1}{a}$ (B) $2a, \frac{2}{a}$ (C) $-a, -\frac{1}{a}$

(D) $1, \frac{1}{a}$ (E) $a^2, \frac{1}{a}$

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	31	36	25	24	14
*A	8	4	10	23	37
B	4	3	3	1	2
C	10	10	13	9	5
D	4	6	8	1	
E	3	1	1	2	2
Total	60	60	60	60	60

Per cent of total group of 300 students answering correctly 27%
Correlation between success on this question and total score on test . . . 53

*Correct answer

Good mathematics achievement questions like this one resemble aptitude test questions in requiring mathematical ingenuity and maturity for their solution. They also require specific content knowledge appropriate to the level of the test, in this case the Intermediate Mathematics Achievement Test offered by the College Entrance Examination Board.

This question is designed to exhibit the general relations between roots and coefficients in the quadratic equation. Able students with clear understandings of these relations will be able to solve this problem by inspection, thus saving time for other parts of the examination. Less able students may be able to solve it by more laborious methods. Students without any knowledge will usually miss or omit it.

The solution depends on the important knowledge that in quadratic equations of the form $x^2 + Bx + C = 0$, C equals the product of the roots and B equals the negative of the sum of the roots. By inspection, one can see that B is written as the negative of the sum of two quantities. The fact that they must be the roots, a and $\frac{1}{a}$, is confirmed by

the fact that their product is C, or 1. A more laborious method involves the use of the quadratic formula, $\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$, where A is the coefficient of x^2 and is equal to 1, B is equal to $-(a + \frac{1}{a})$, and C is equal to 1. A third and also laborious method involves factoring the left member of the equation and setting each factor equal to zero.

The discrimination power of the question can be observed by noting the ability level of the students selecting the correct answer. A progression of students choosing (A), the right answer, can be seen from 8 in the lowest ability group to 37 in the highest ability group. Choice (C) attracted 47 students, probably because of a mixup in sign in selecting the sum of the roots. There is some reason to believe that students who chose (B) used the quadratic formula above and failed to divide by 2 at the end.

Statistical Analysis

	Students Classified by Total Test Score				
RESPONSES	LOWEST FIFTH	NEXT LOWEST FIFTH	MIDDLE FIFTH	NEXT HIGHEST FIFTH	HIGHEST FIFTH
Omit	11	6		1	
A	9	4	7	2	1
B	4	8	9	13	14
C	2	7	3	7	1
*D	11	19	30	24	37
E	21	15	11	13	6
Total	58	59	60	60	59

Per cent of total group of 296 students answering correctly 41%
Correlation between success on this question and total score on test . . .38

(Discrepancies in total numbers in each fifth are caused by drop-out of students not completing the test.)

*Correct answer

Question 22

The question below is followed by two statements, labeled (1) and (2), in which certain data are given. In this question you do not actually have to compute an answer, but rather you have to decide whether the data given in the statements are sufficient for answering the question. Using the data given in the statements plus your knowledge of mathematics and everyday facts (such as the number of days in July), you are to select answer

(A) if statement (1) ALONE is sufficient but statement (2) alone is not sufficient to answer the question asked,

(B) if statement (2) ALONE is sufficient but statement (1) alone is not sufficient to answer the question asked,

(C) if both statements (1) and (2) TOGETHER are sufficient to answer the question asked, but NEITHER statement ALONE is sufficient,

(D) if EACH statement is sufficient by itself to answer the question asked,

(E) if statement (1) and (2) TOGETHER are NOT sufficient to answer the question asked and additional data specific to the problem are needed.

If x is a whole number, is it a two-digit number?

(1) x^2 is a three-digit number.

(2) 10x is a three-digit number.

One of the abilities which has been receiving increasing emphasis from the elementary school through college is that of judging the relevancy of data in the solution of problems in mathematics, science, and social studies. Measurement of the extent to which a student has developed this ability near the end of the high school years is believed to be important in predicting scholastic success in college. To accomplish such measurement the type of test question presented here has been designed, validated by research, and is now in use in the College Entrance Examination Board's Scholastic Aptitude Test. This type of question is used in arithmetic, algebra, and geometry. It shifts the emphasis from rote, manipulative skills to higher level judgments and reasoning.

This question shows the very great versatility of this type, for it requires little factual knowledge and the simplest aspects of elementary algebra, but does require a considerable degree of numerical judgment.

Fact 1 alone is sufficient because the square root of any three-digit square is a two-digit number. Fact 2 alone is sufficient because, whenever a three-digit multiple of 10 is divided by 10, the result is a two-digit number. The correct answer is therefore (D).

This question and its analysis provide an interesting insight into the functioning of this type of question. When the correct answer is (D), there is a sense in which the question becomes two questions, since two separate sufficiencies must be determined. In this case, when the sufficiency of Fact 1 is considerably more difficult to determine than the sufficiency of Fact 2, a certain number of moderately able people can be expected to choose (B) as the answer. One of the reasons that may have made the sufficiency of Fact 1 difficult for them to see is that although the square root of any three-digit number is always a two-digit number, the square of a two-digit number is not necessarily a three-digit number. The latter point may have led some to conclude erroneously that Fact 1 was not sufficient. Fact 2, on the other hand, can more easily be seen as sufficient because only a two-digit number can be multiplied by 10 to produce a three-digit number. The higher ability of the 121 selecting (D) is sufficient to give the question a good index of discrimination. The 20 choosing (C) apparently failed to realize that each of the two facts is independently sufficient.